

PERSONALIZACION DE AVATARES HUMANOS MEDIANTE MODELOS DE DEEP LEARNING

Autor: Javier Serrano del Amo

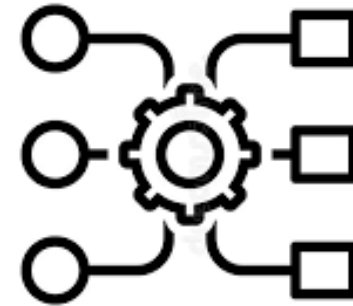
Tutor: Jorge Félix López Moreno

Co-Tutor: Dan Casas Guix



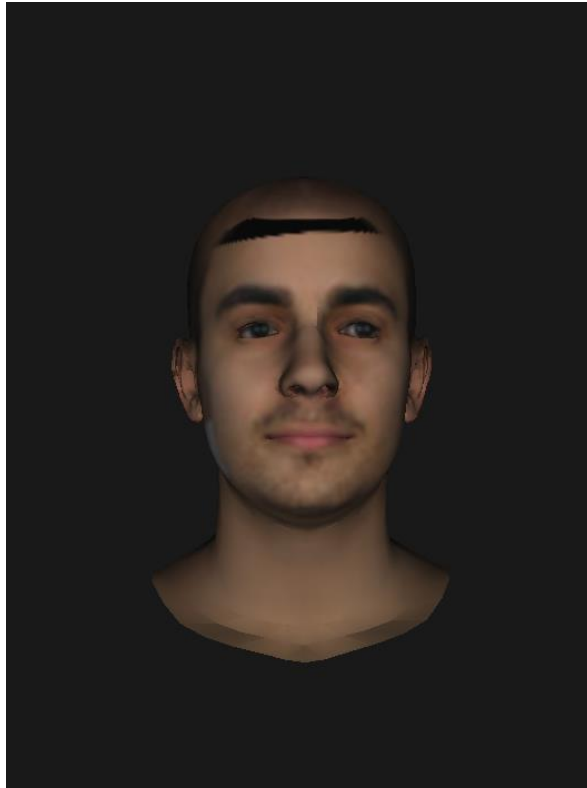
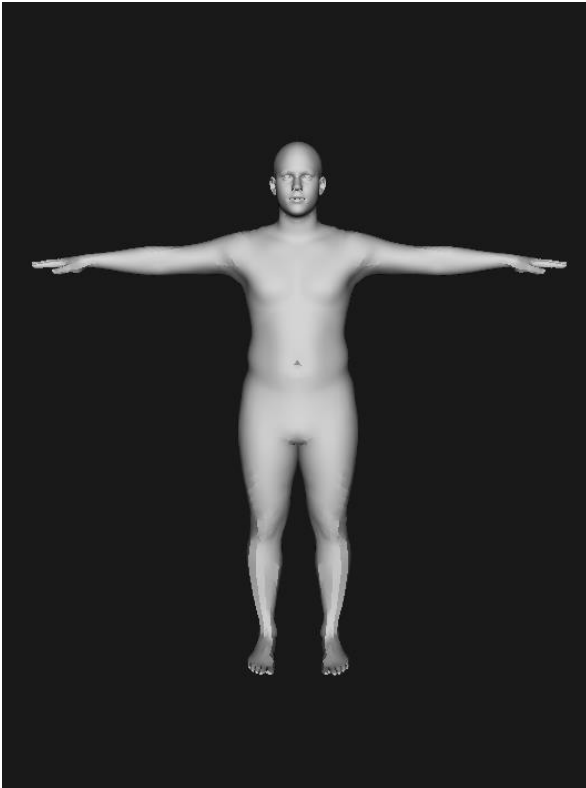
INTRODUCCIÓN

- Gran interés en el mundo digital.
- Proceso lento y gastos de recursos.
- Aparición de algoritmos más sofisticados.
- Necesidad de automatizar los procesos.
- Mínimos datos de entrada.
- Sencillez de uso y edición de cara al usuario final.



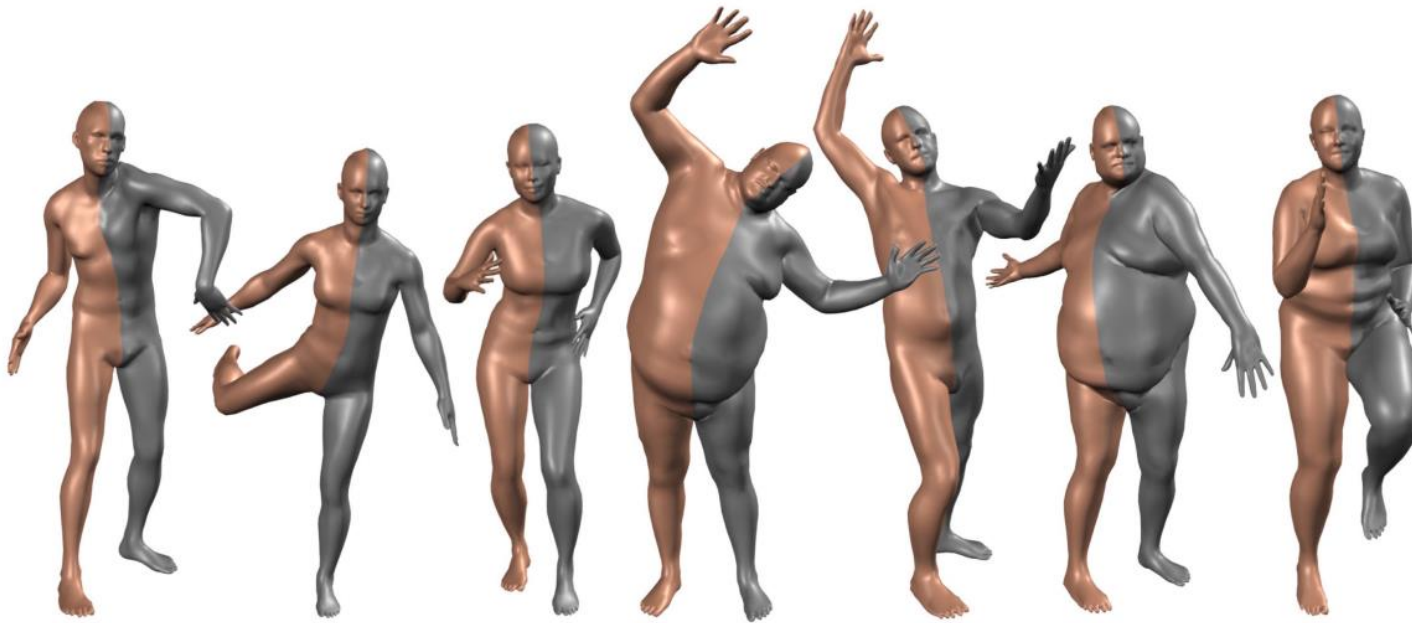
¿CÓMO?

- Unión de modelos paramétricos de vanguardia SMPL-X y DECA.



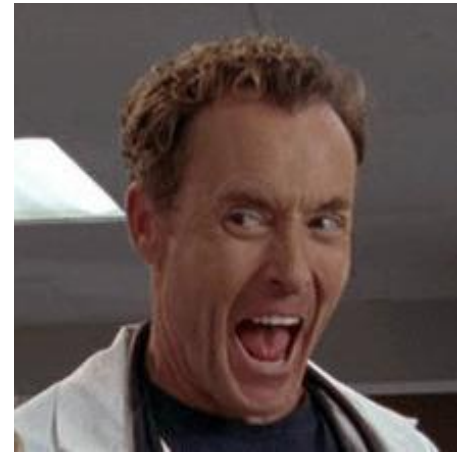
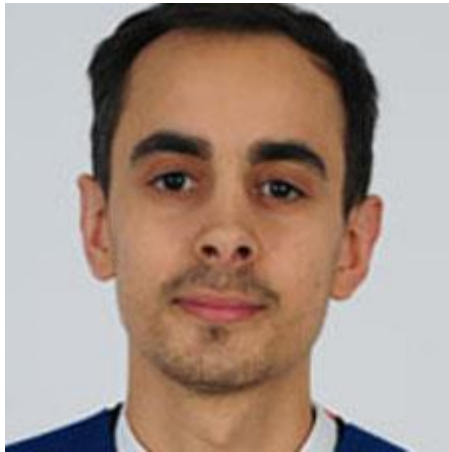
SMPL (Skinned Multi-Person Linear Model)

- Entrenado mediante el CAESAR *dataset* (2000 escáneres humanos).
- 10 parámetros.
- Cuerpos de 6890 vértices.

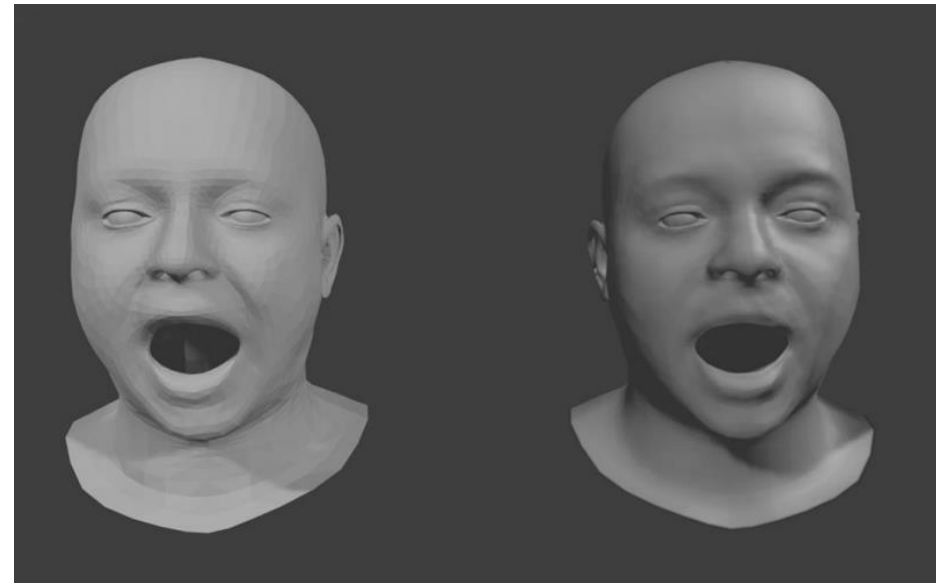


DECA (Detailed Expression Capture and Animation)

- A partir de imágenes de 224x224 píxeles.
- Distinción de rasgos de identidad y expresión.

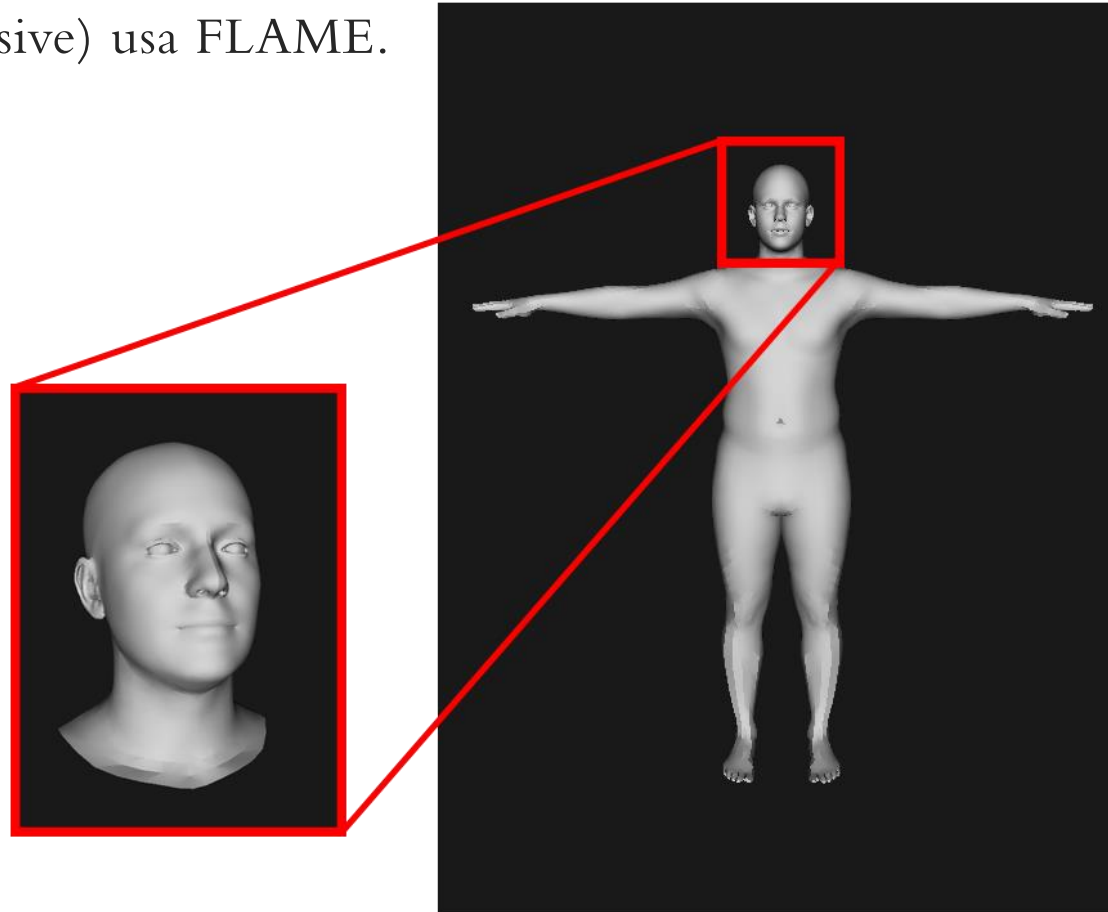


- Texturización automática.
- Construido sobre FLAME(Faces Learned with an Articulated Model and Expressions), abstrayendo sus parámetros.
- 5023 vértices.



¿POR QUÉ SMPL-X EN VEZ DE SMPL?

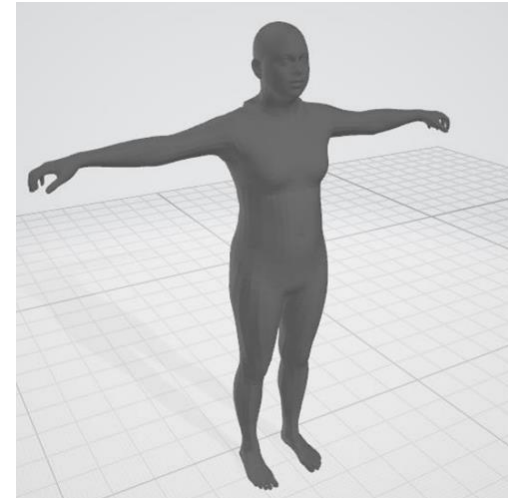
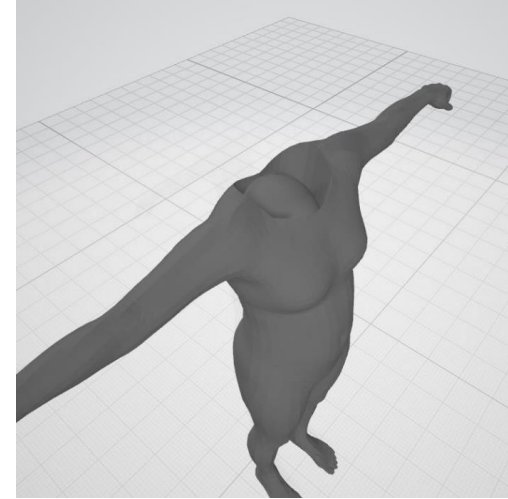
- SMPL-X (SMPL eXpressive) usa FLAME.
- 10475 vértices.
- Punto de unión.



ESTIMACIÓN DE GEOMETRÍA

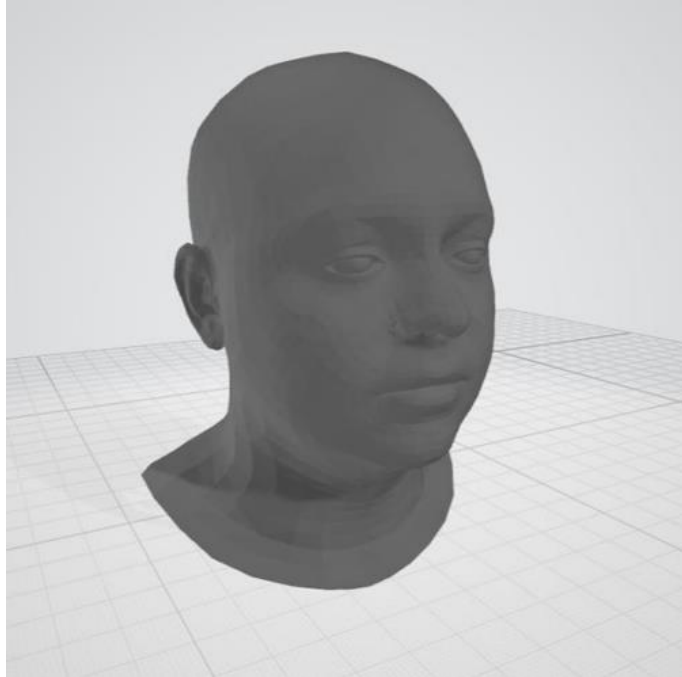
FLUJO DE TRABAJO DE GEOMETRÍA

- Punto de unión.
- Posicionamiento de cabeza.
- *Offsets* con FLAME neutro.



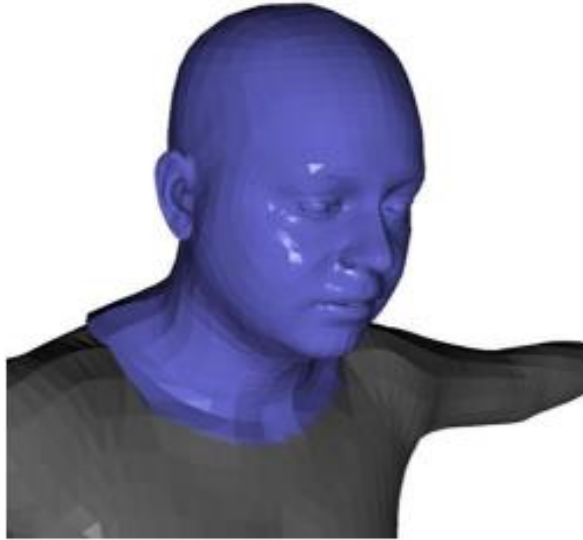
$$O(m_{\text{malla } 1}, m_{\text{malla } 2}) = \\ = \{v_i | v_i = v_i^{m_{\text{malla } 2}} - v_i^{m_{\text{malla } 1}}, \forall i = 1, 2, \dots, 5023\}$$

- Abertura en boca.
- Las cabezas no tienen el mismo pivote.

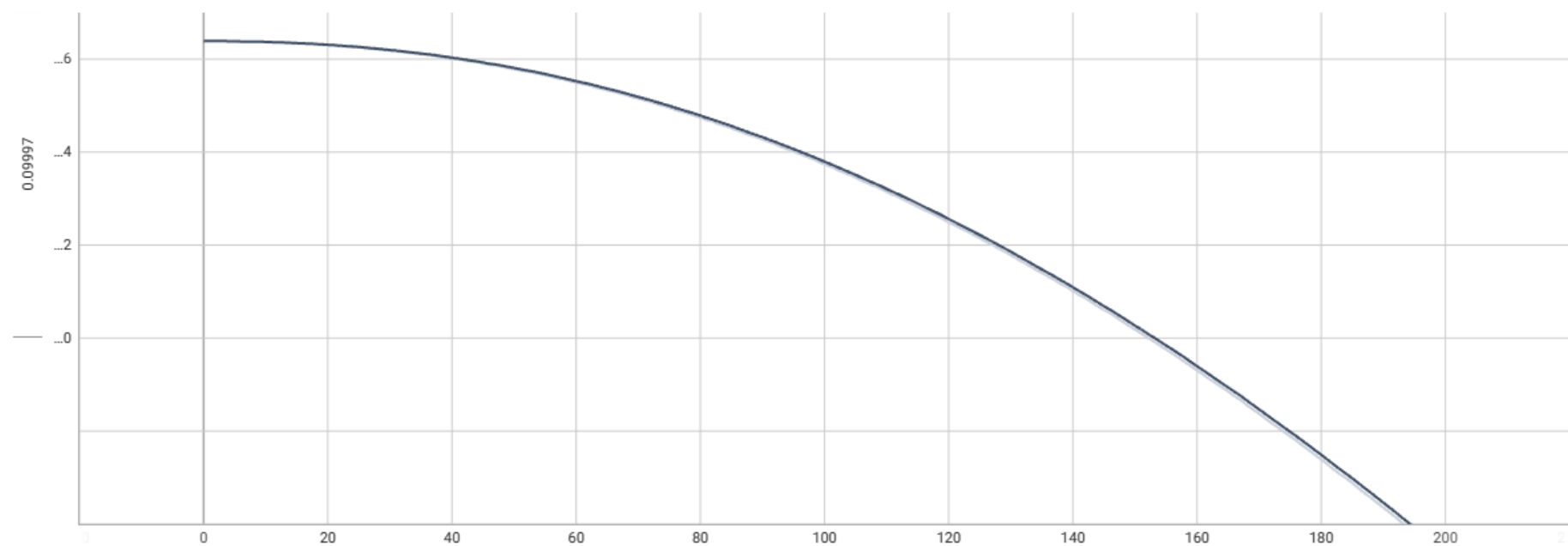


- Creación de optimizador.
 - Optimización de las coordenadas xyz.
 - Alineación inicial aproximada basada en la media de los vértices.
 - Uso de optimizador Adam.

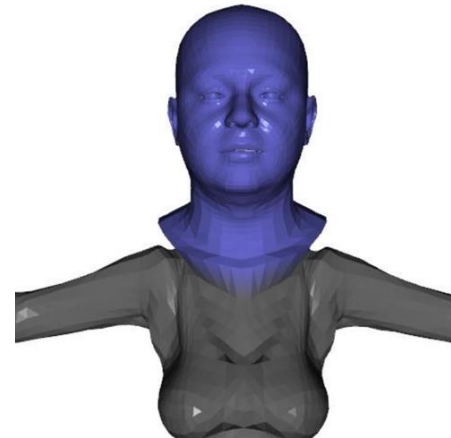
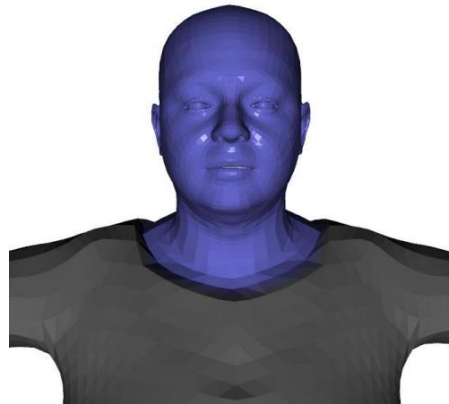
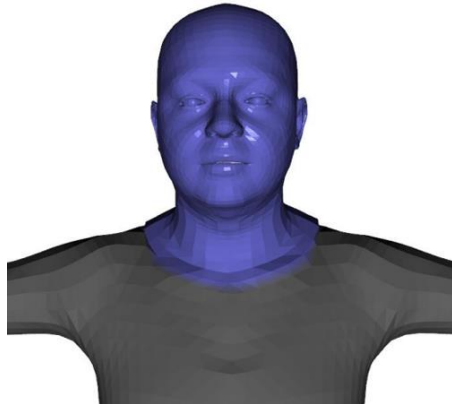
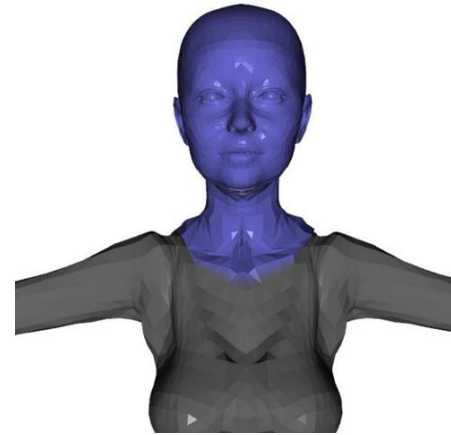
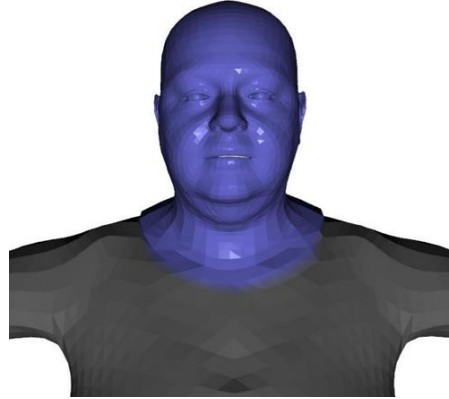
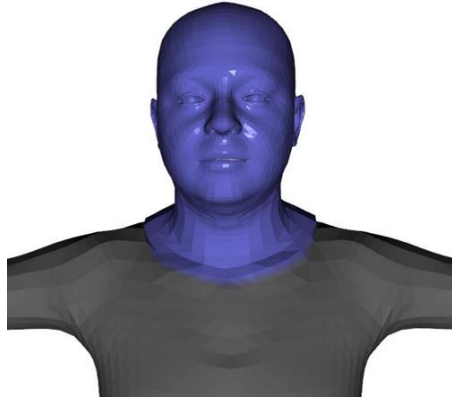
- Función de pérdida. $Loss = \sum_{i=1}^{5023} (v_{i,malla_1} - v_{i,malla_2})^2$



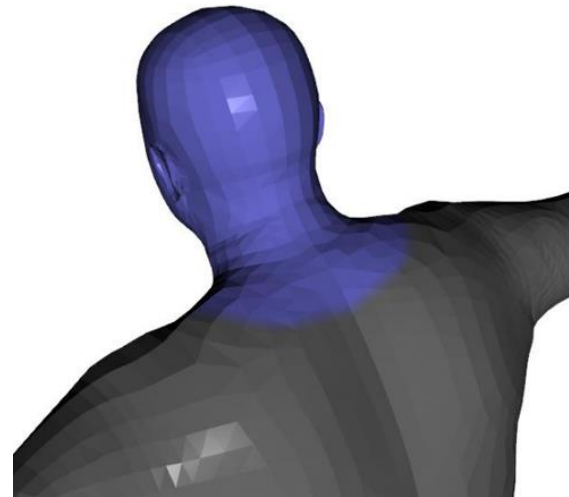
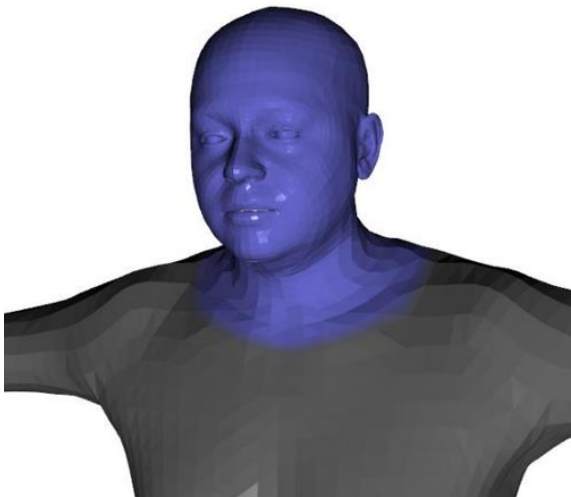
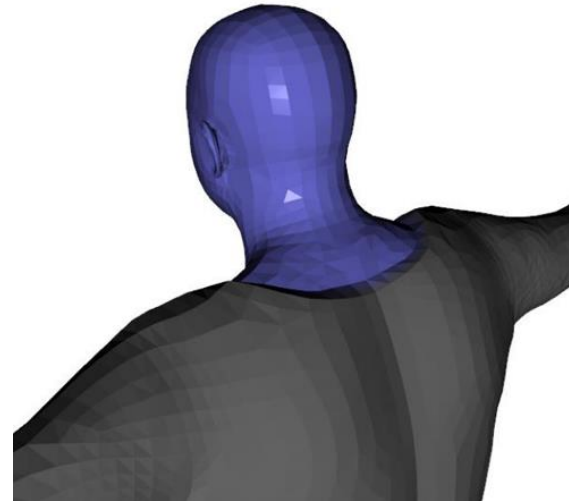
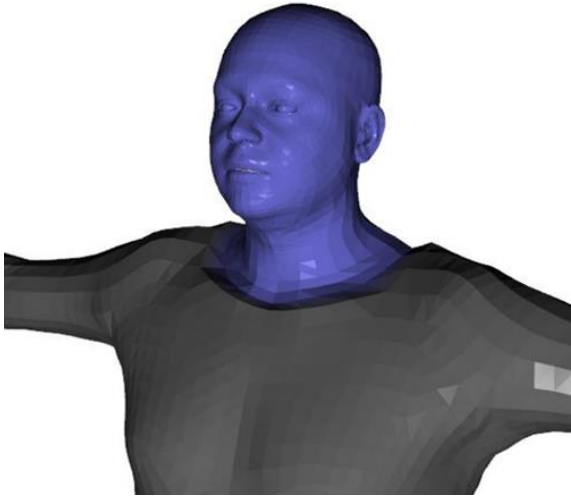
Alignment loss



- Uso de cuerpos personalizados. No neutros.
 - Modificación de cabeza SMPL-X.
 - Nuevos Offsets entre cabeza SMPL-X personalizada y SMPL-X neutro.
 - Offsets para los modelos con expresión.

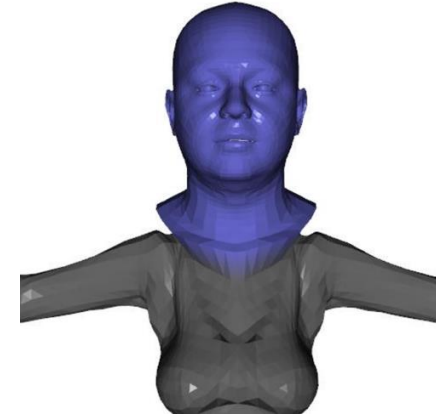


- Suavizado de clavícula/cuello.



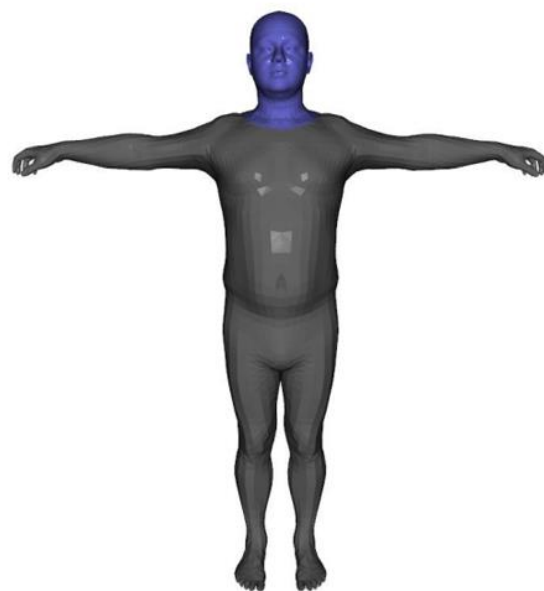
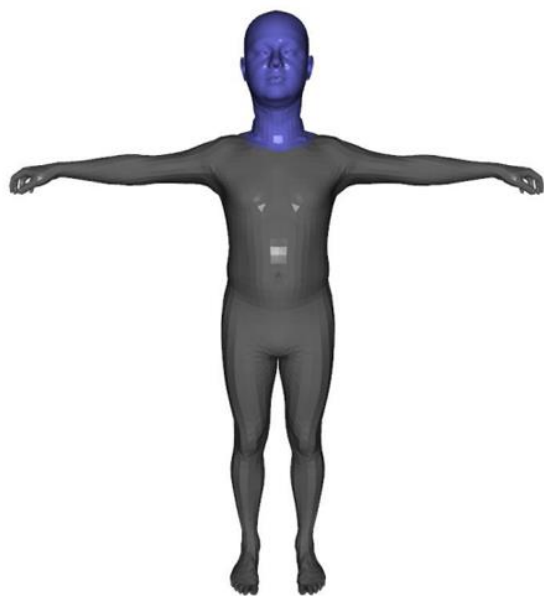
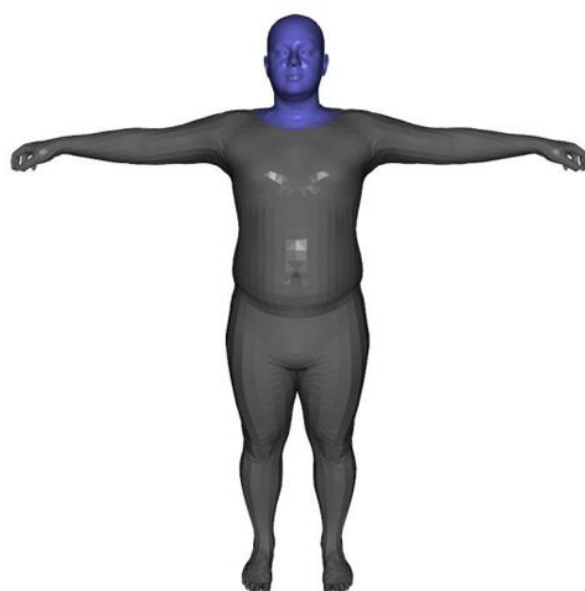
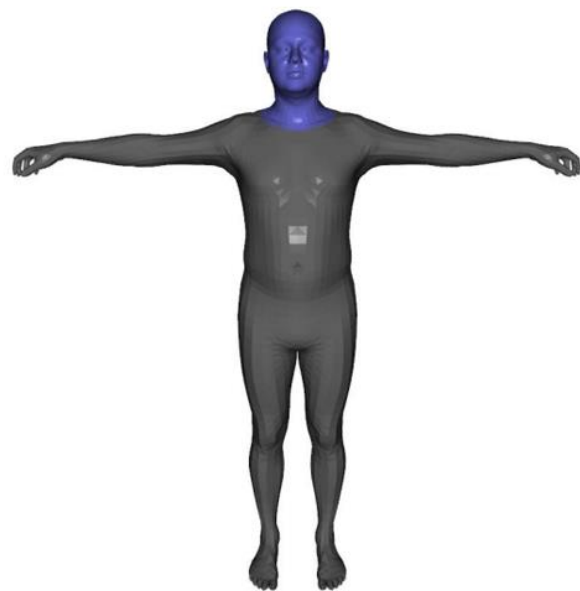


- Creación de cuerpo automática basada en cabeza generada por DECA.
 - La cabeza no se ajusta al cuerpo.
 - El usuario debe escoger el cuerpo de forma correcta.

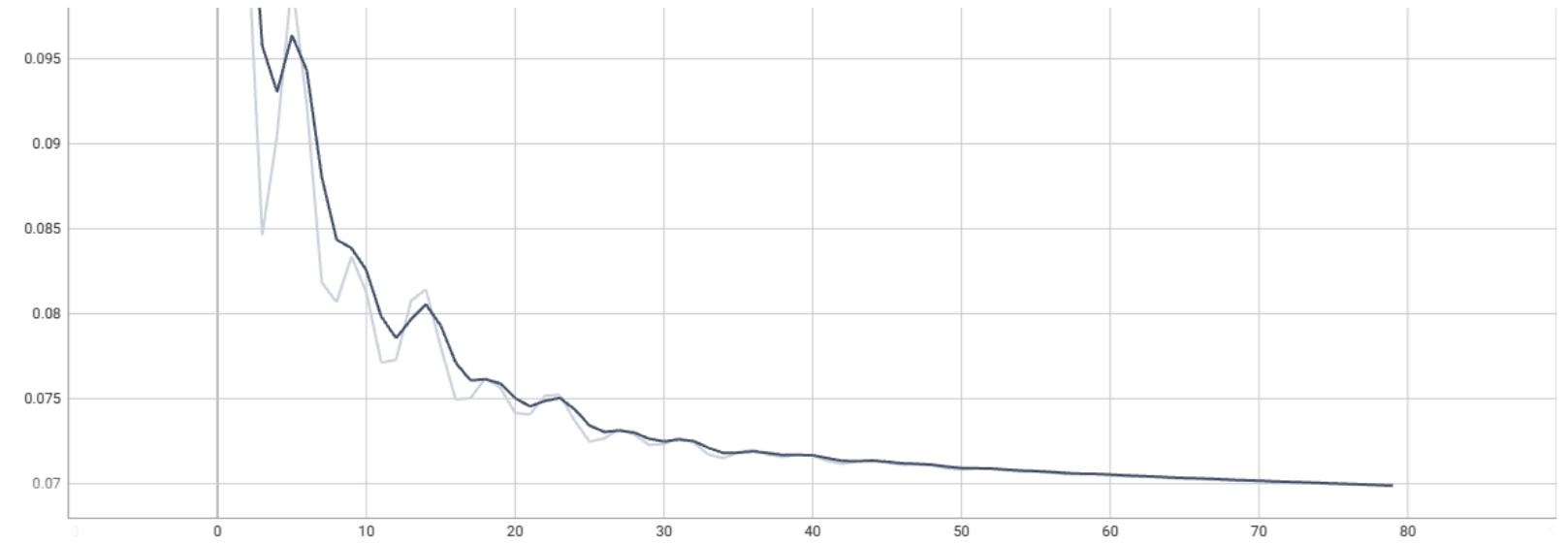


- Inferencia de cuerpo mediante optimizador.
 - 10 parámetros optimizados.
 - Optimizador Adam.
 - Reducción de función de pérdida.

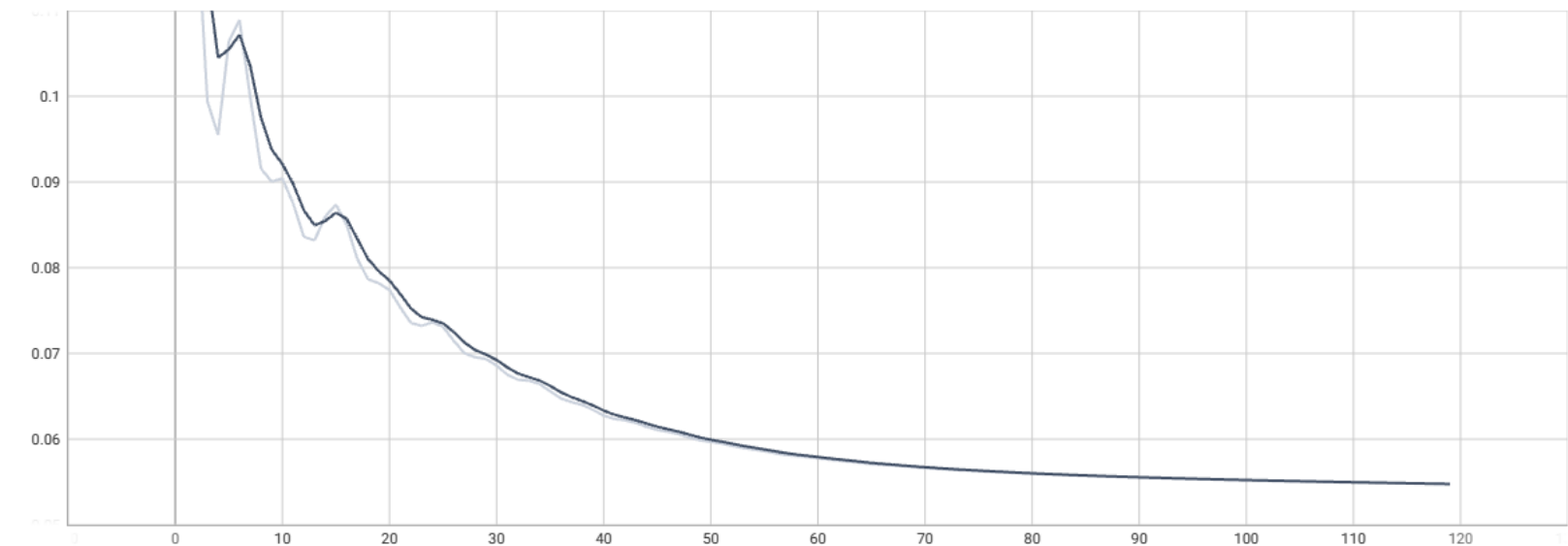
$$Loss = \sum_{i=1}^{5023} (v_{i, \text{malla}_1} - v_{i, \text{malla}_2})^2$$



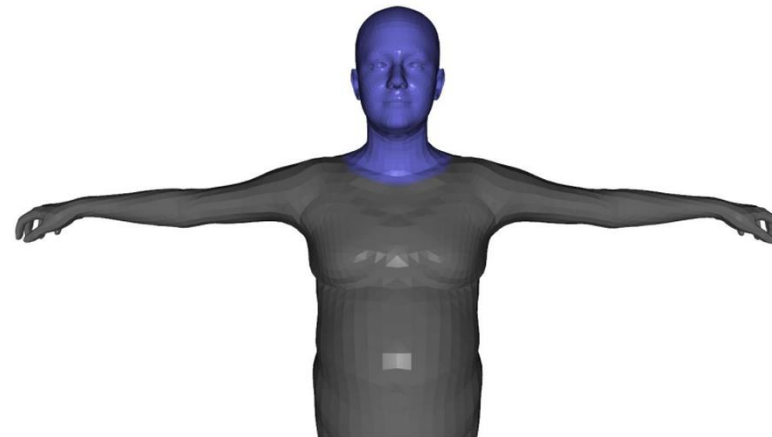
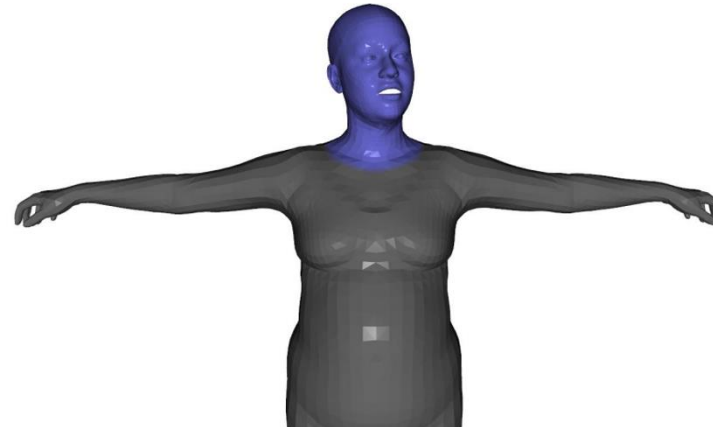
Inference loss

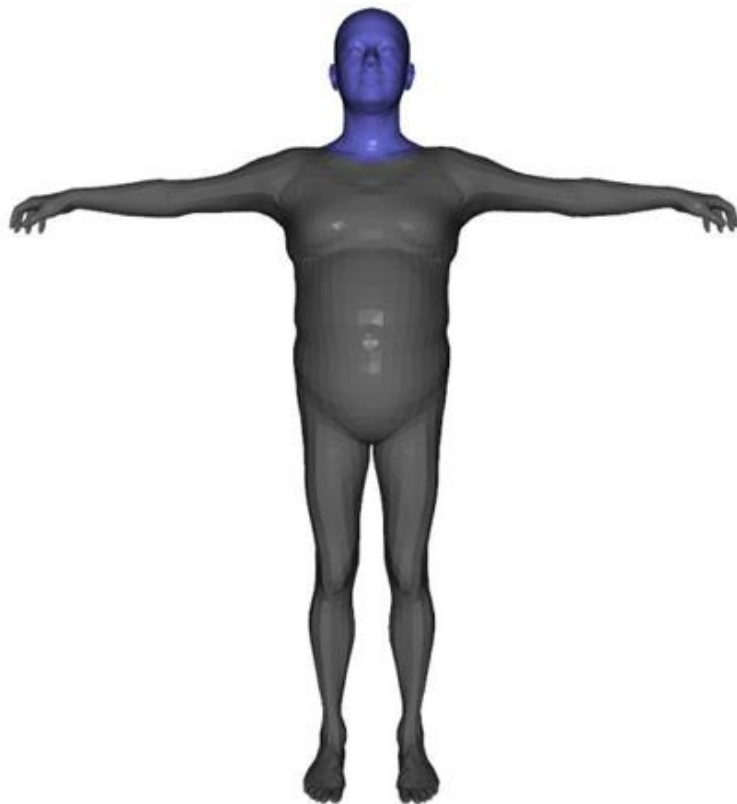
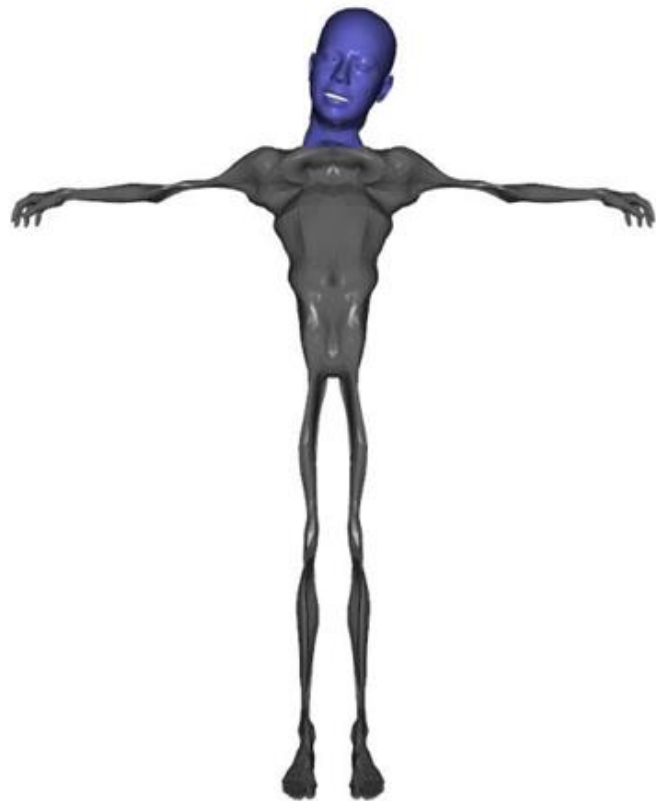


Inference loss

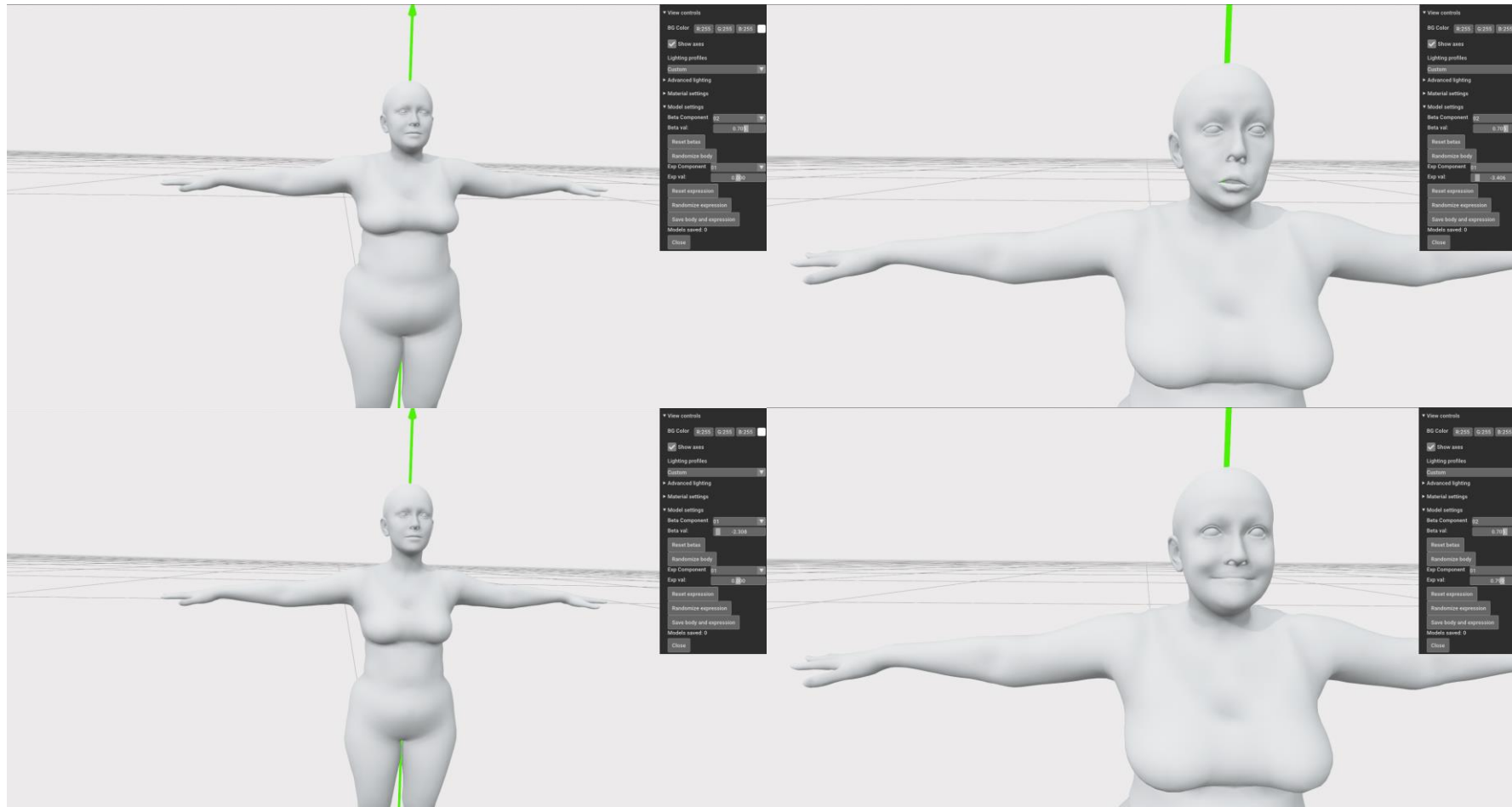


- Corrección de pose generada por DECA.
 - DECA usa un espacio latente con diversos propósitos.





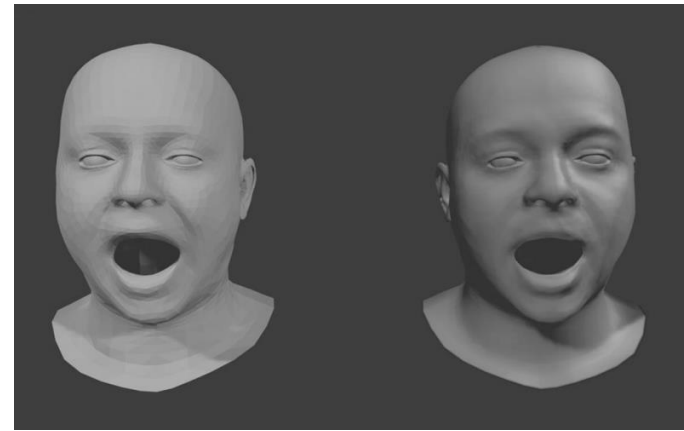
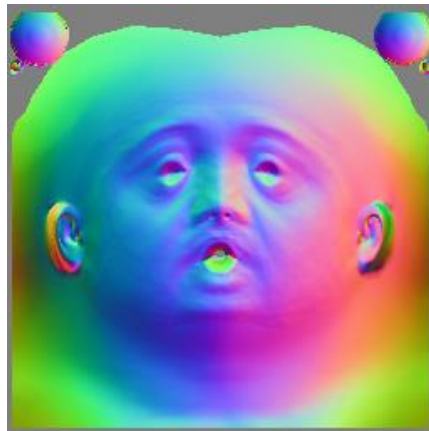
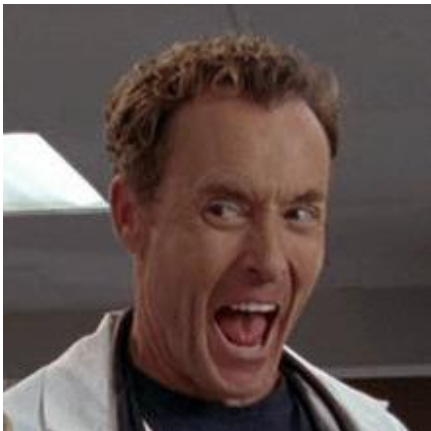
- Creación de interfaz gráfica de usuario.
 - Necesidad de ver los efectos de cada parámetro sobre el cuerpo.
 - Creación de cuerpo en tiempo real.
 - Creación de expresiones personalizadas.



ESTIMACIÓN DE TEXTURAS

FLUJO DE TRABAJO DE TEXTURAS

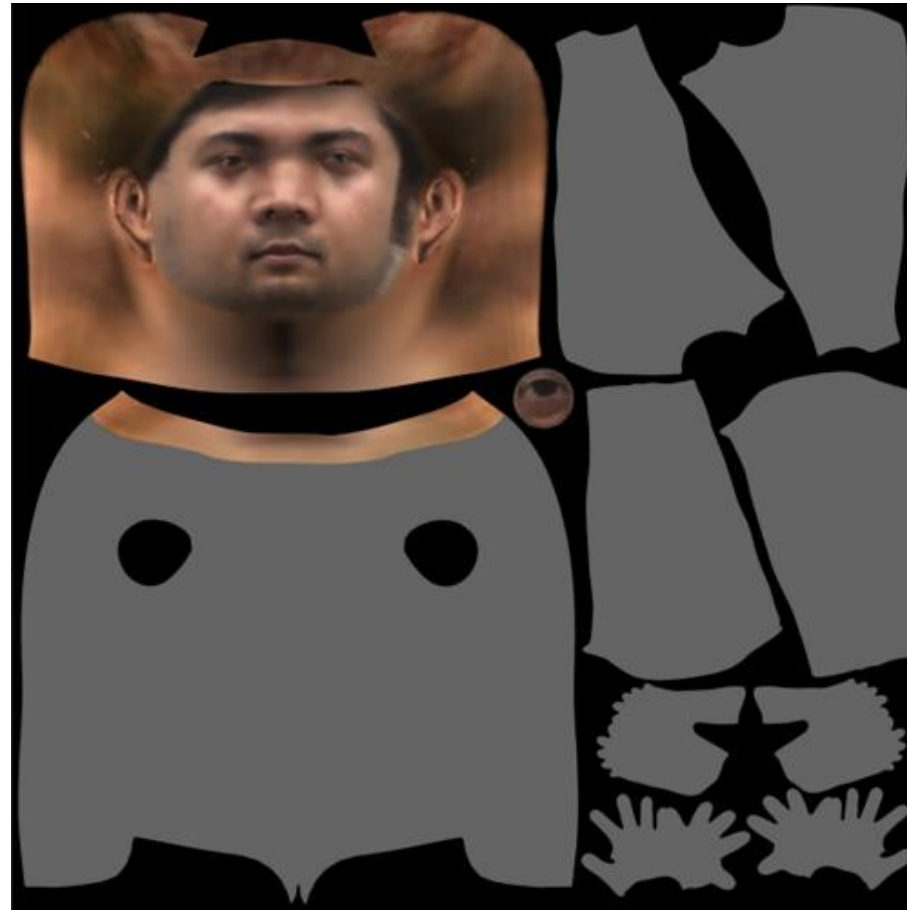
- DECA es capaz de generar varias texturas.



- Corrección de textura por corrección de pose.
 - 236 valores en el espacio latente.
 - Valores de la cámara influenciados por pose.



- Texturización de cabeza en cuerpo SMPL-X.
 - Sustitución de coordenadas UV.
 - Uso de proyecto externo.



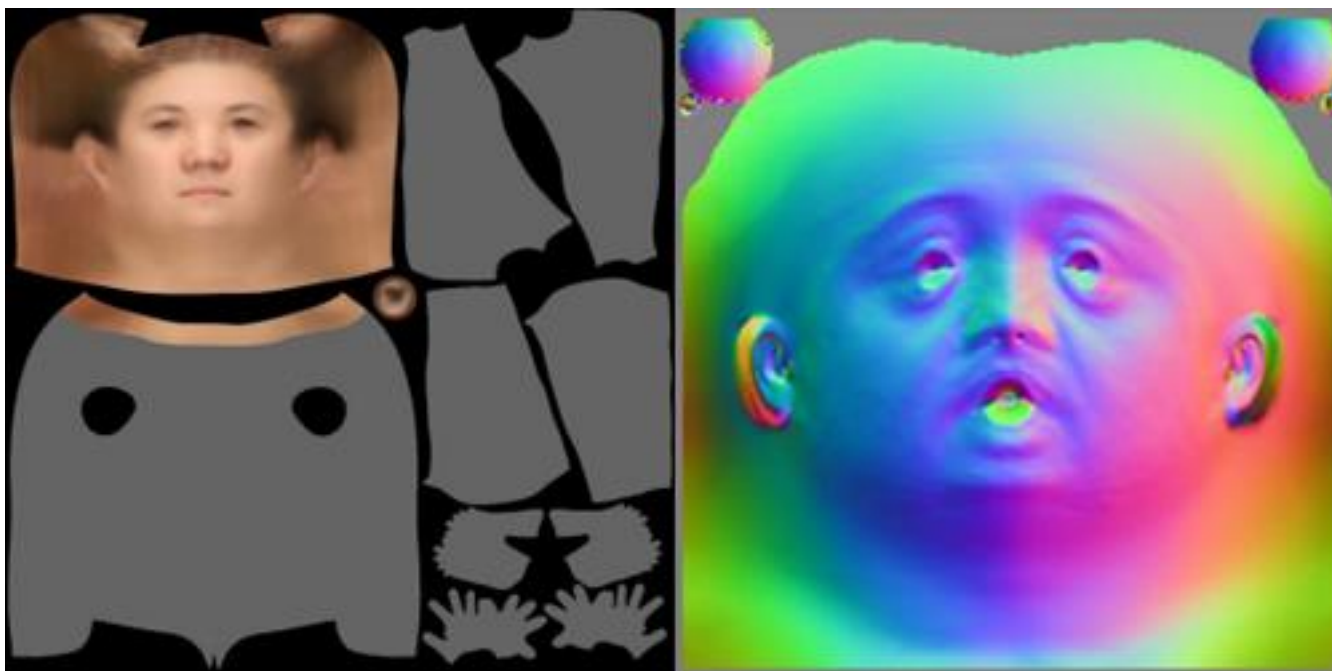
- **Inconvenientes :**

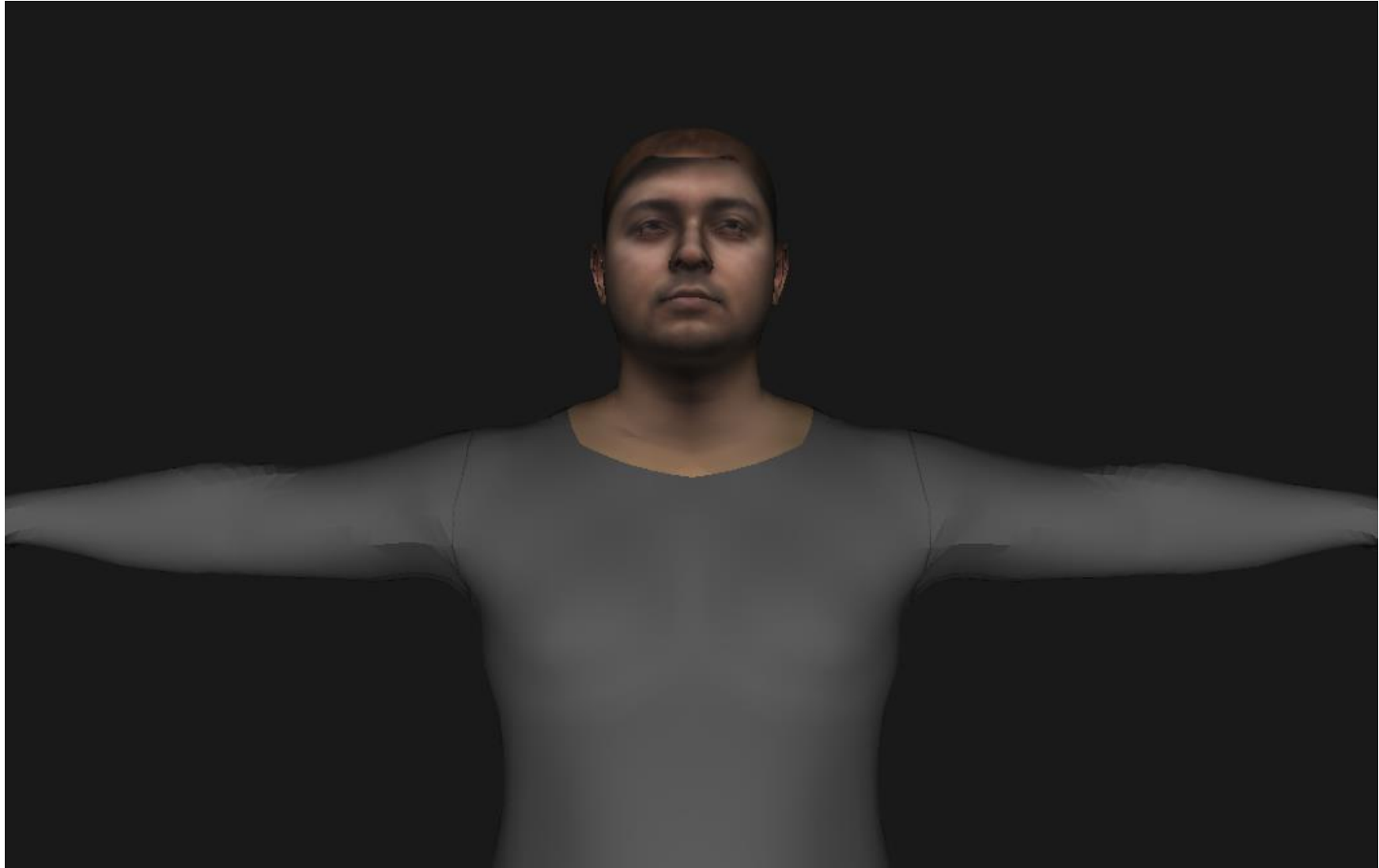
- Tiempos de ejecución muy largos (1 hora por textura).
- Warping.
- Escalado de texturas.

- **Soluciones :**

- Mejor gestión de la resolución de texturas.
- Mejora de código.
- Paralelización mediante joblib.
- Tiempo de ejecución resultante (3:30 minutos para las dos texturas).

- Gracias al uso de las correspondencias entre vértices se puede...





- *Inpainting*.
 - Rellenado de zonas oscuras.
 - Basado en la vecindad de los píxeles cercanos.
 - Herramienta manual.

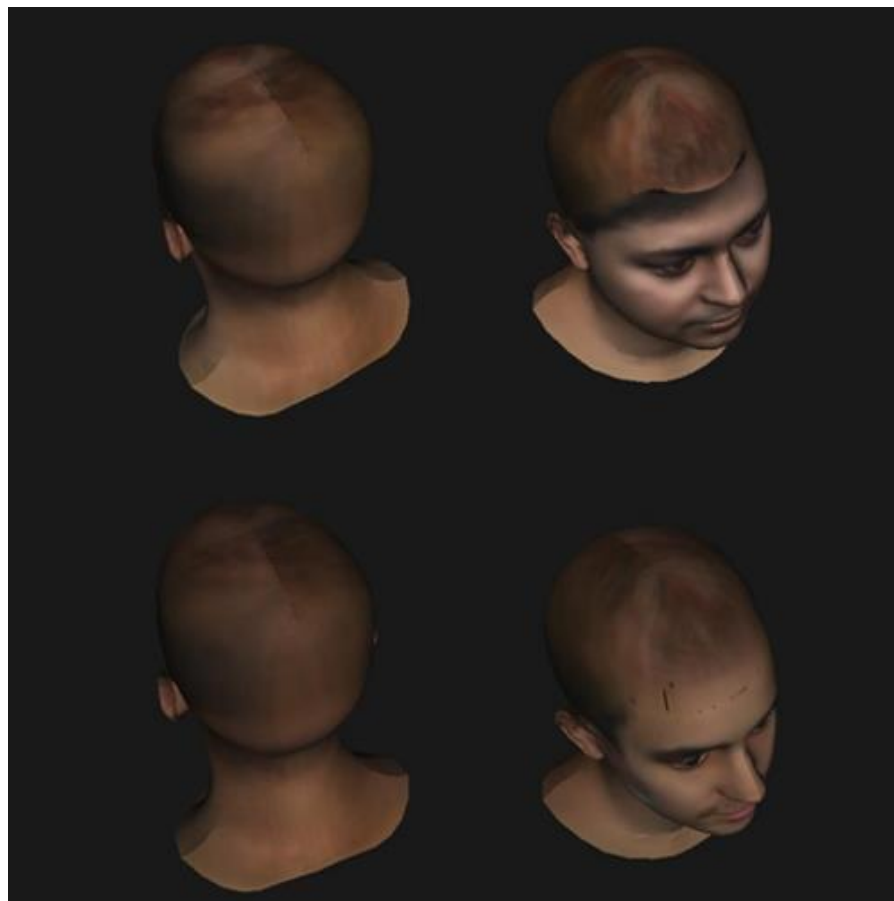
$$factor = \sum_{x=1}^n \sum_{y=1}^m (0,2126 \cdot B_{x,y} + 0,7152 \cdot G_{x,y} + 0,0722 \cdot R_{x,y})$$





- Corrección de efecto costura.

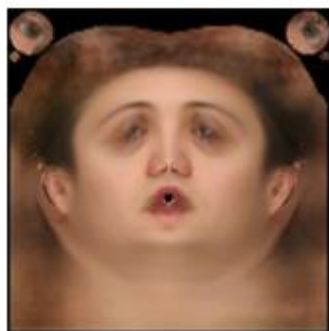




- Alineación de ojos mediante red neuronal.



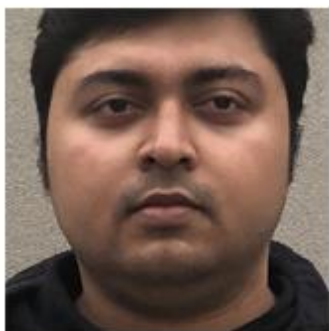
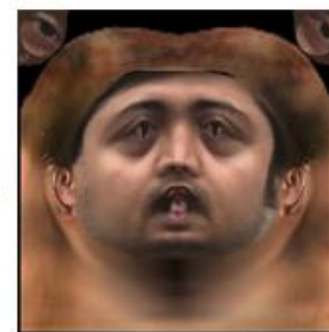
- ¿Mala predicción de *landmarks*?
 - Modificación del entrenamiento para hacer *overfitting* en base a la función de error de *landmarks*.
 - Modificación de imagen de entrada.
-
- Mismos resultados, ¿qué ocurre? ¿Cómo genera DECA el albedo?
 - Codificación de textura(intermedia) en 50 parámetros.



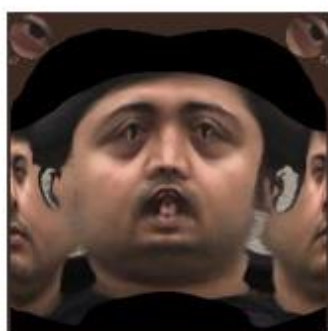
*



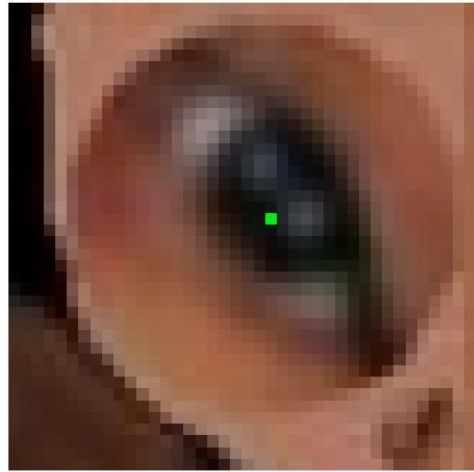
=



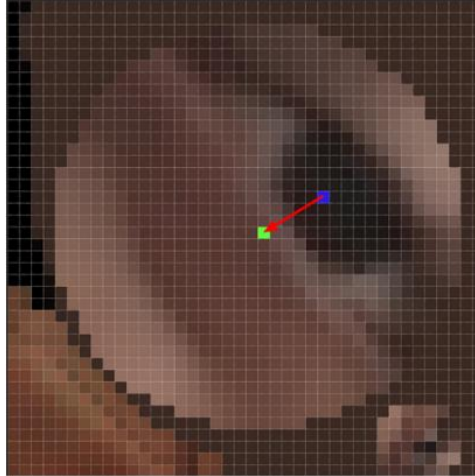
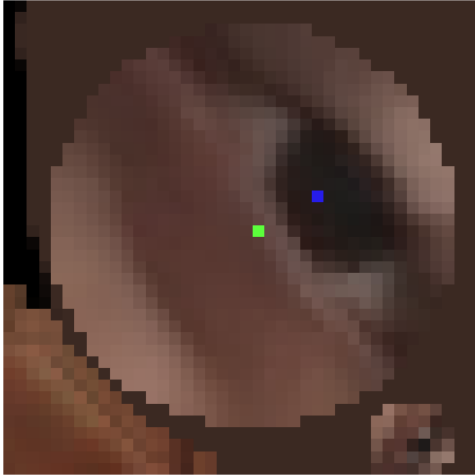
+ renderizador =



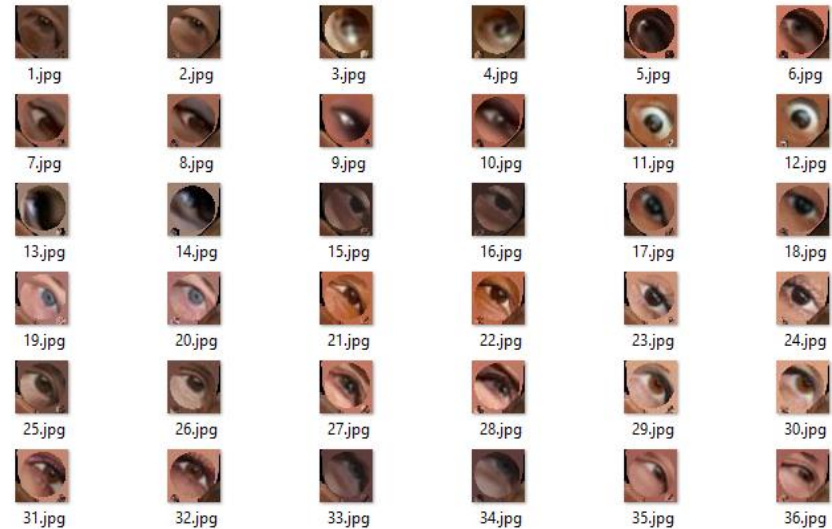




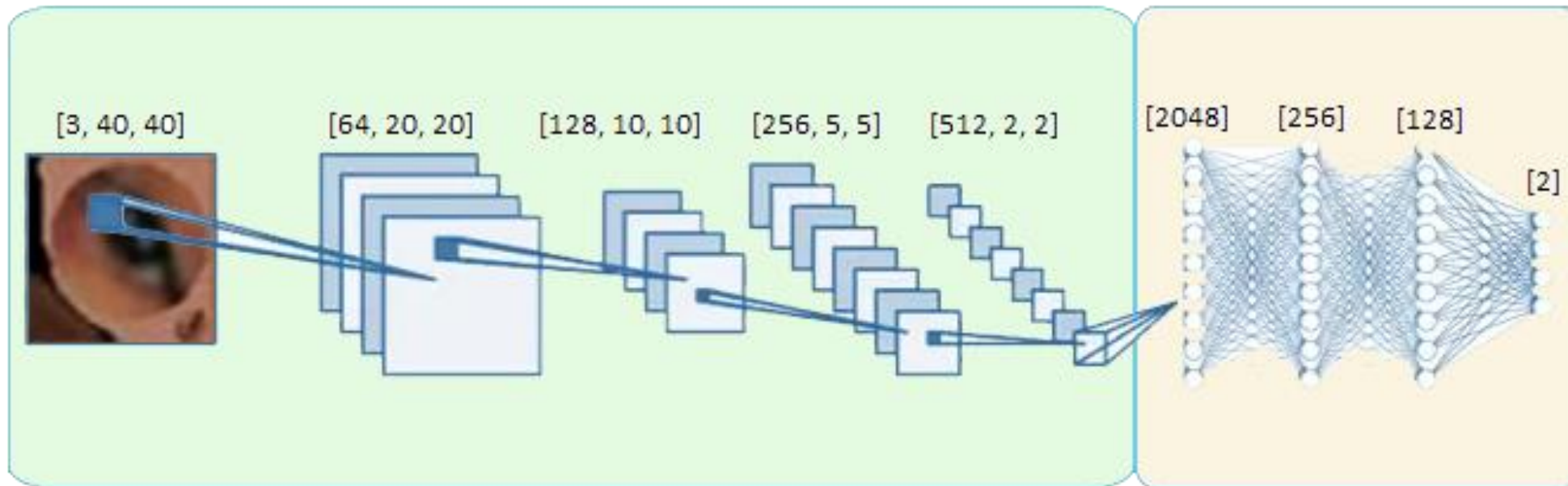
- Red neuronal que haga una predicción del centro deseado.



- *Dataset* personalizado (82 ojos).
- 40x40 píxeles.
- ¿Arquitectura de red neuronal?



- Capas convolucionales para la extracción de patrones y estructuras.
- MLP completamente conectada para predicción de píxeles.

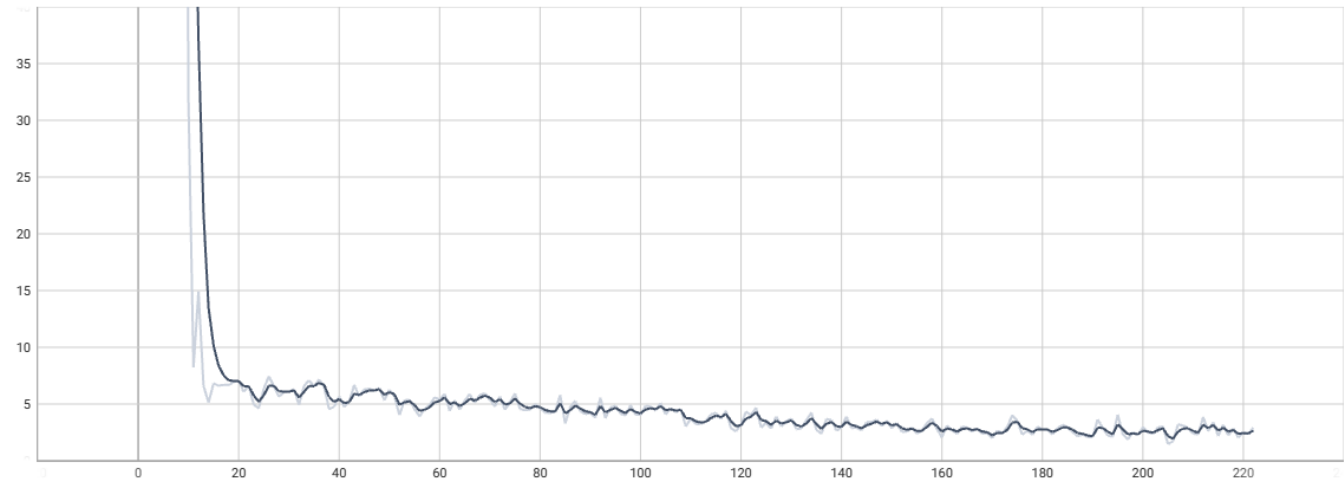


- Función de pérdida MSE(Mean Squared Error).

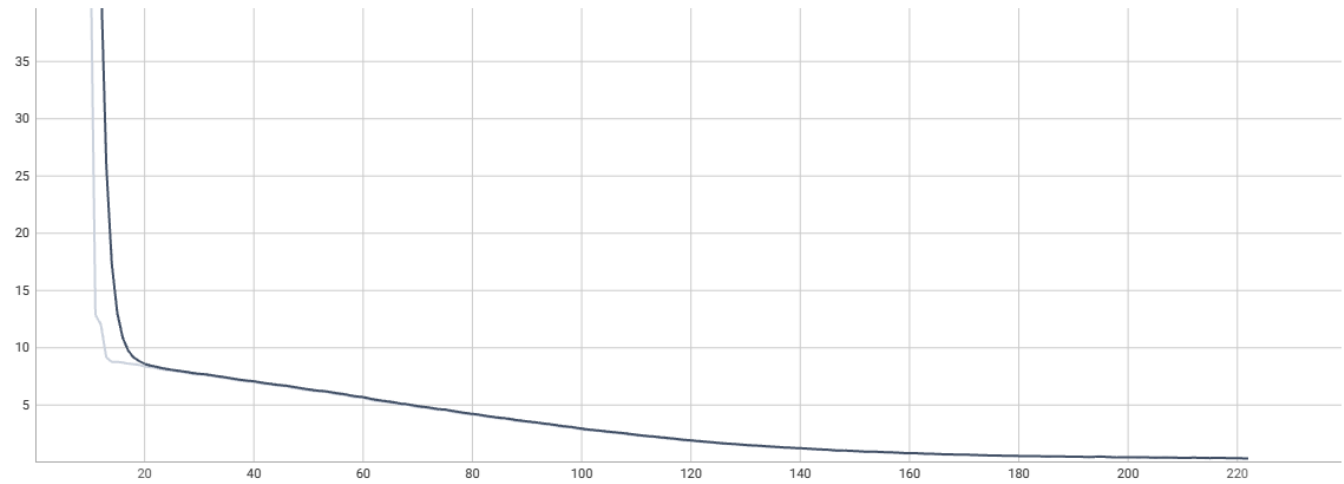
$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

- Optimizador Adam.
- *Learning rate* de 0,0001.
- División del *dataset* en entrenamiento(80%), validación(10%) y testing(10%).

Training loss

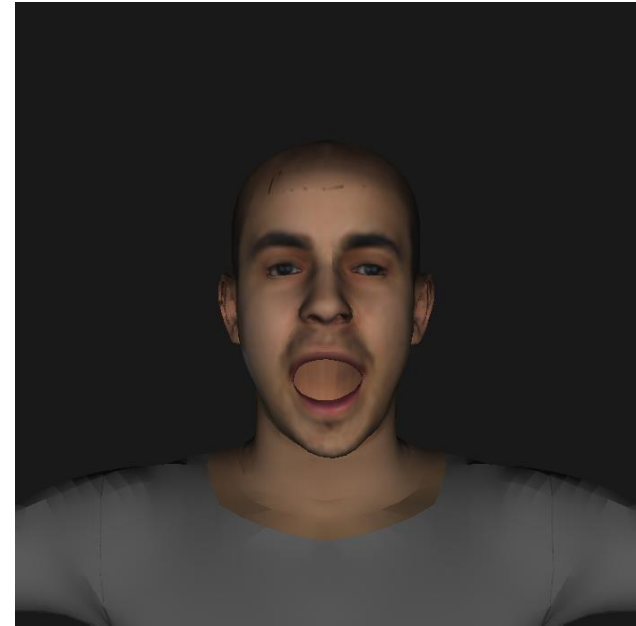
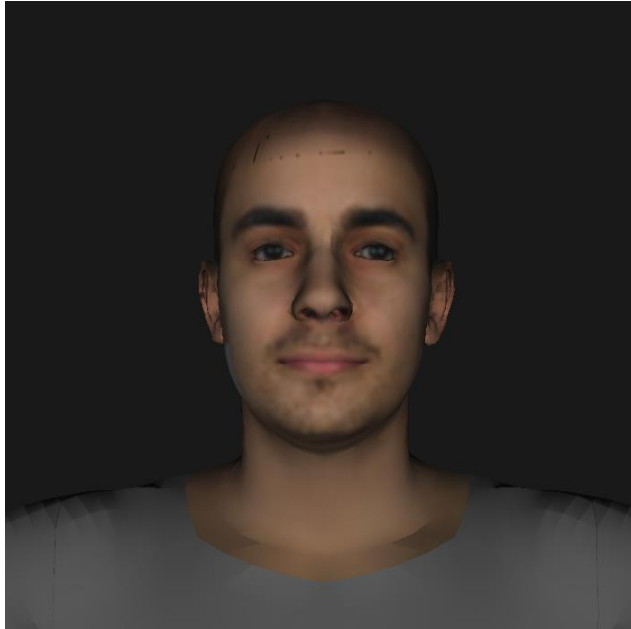


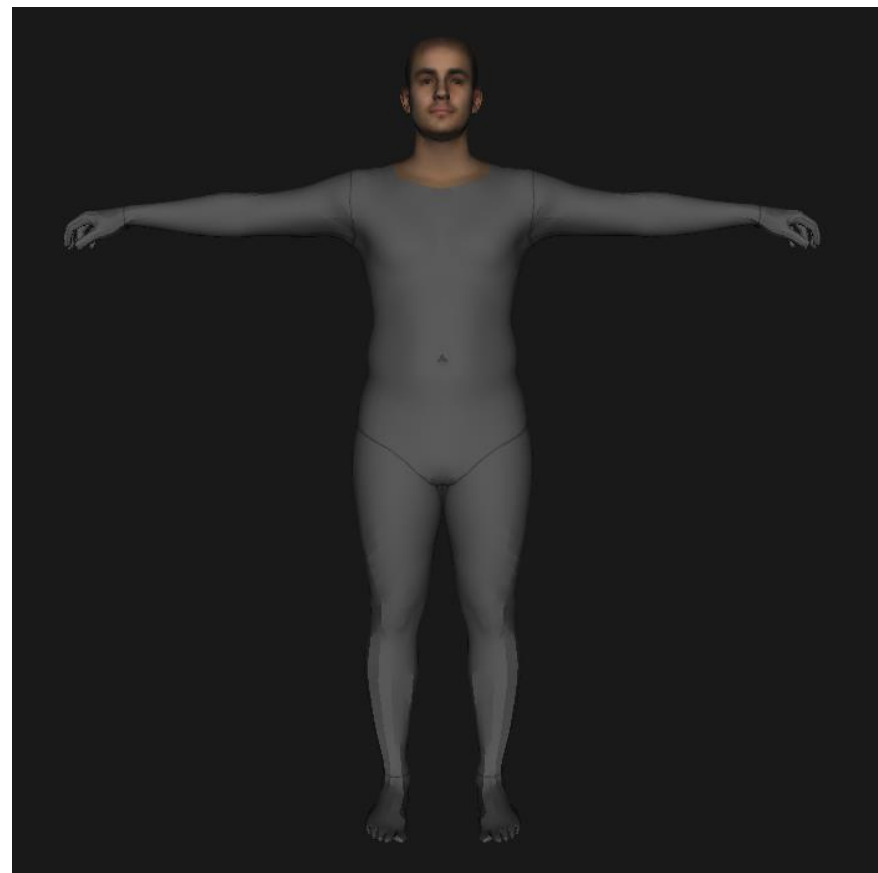
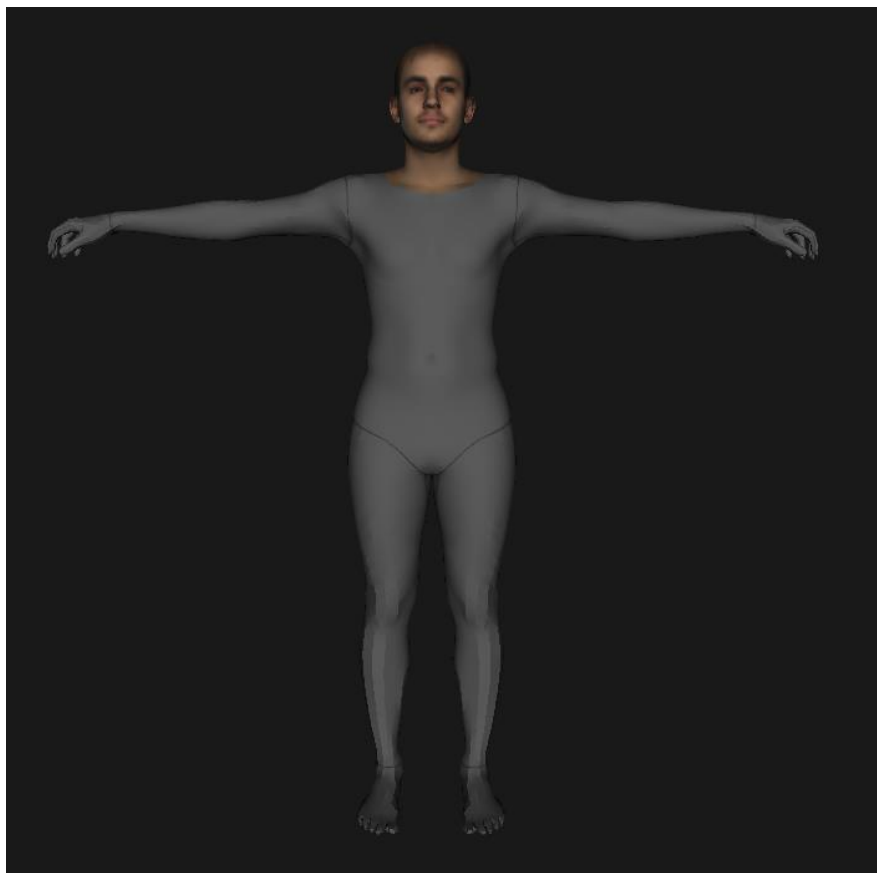
Evaluation loss





RESULTADOS





TRABAJOS FUTUROS

- Mejoras en la eficiencia computacional.
- Mejoras en la calidad de la texturización.
- Texturización de cuerpo.
- Extensión a multivista.



GRACIAS POR
ESCUCHARME

¿ALGUNA
PREGUNTA?